GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L24402, doi:10.1029/2006GL027941, 2006

Nonlinear statistics reveals stronger ties between ENSO and the tropical hydrological cycle

Shiraj Khan,[1,2] Auroop R. Ganguly,[1] Sharba Bandyopadhyay,[3] Sunil Saigal,[2] David J. Erickson III,[4] Vladimir Protopopescu,[1] and George Ostrouchov[4]

Received 12 September 2006; revised 27 October 2006; accepted 6 November 2006; published 20 December 2006.

[1] Cross-spectrum analysis based on linear correlations in the time domain suggested a coupling between large river flows and the El Niño-Southern Oscillation (ENSO) cycle. A nonlinear measure based on mutual information (MI) reveals extrabasinal connections between ENSO and river flows in the tropics and subtropics, that are 20–70% higher than those suggested so far by linear correlations. The enhanced dependence observed for the Nile, Amazon, Congo, Paraná, and Ganges rivers, which affect large, densely populated regions of the world, has significant impacts on inter-annual river flow predictabilities and, hence, on water resources and agricultural planning. **Citation:** Khan, S., A. R. Ganguly, S. Bandyopadhyay, S. Saigal, D. J. Erickson III, V. Protopopescu, and G. Ostrouchov (2006), Nonlinear statistics reveals stronger ties between ENSO and the tropical hydrological cycle, *Geophys. Res. Lett.*, *33*, L24402, doi:10.1029/2006GL027941.

1. Introduction

[2] ENSO events impact regional precipitation in the tropics and subtropics, ultimately causing inter-annual variability in river flows. The ocean-atmosphere-land interactions are complex and far from being completely understood and accurately modeled. A slight disturbance in these interactions would usually result in sometimes surprising distant correlations and climate patterns. Analyses of the rainfall anomalies during the warm (El Niño) and cold (La Niña) episodes of ENSO suggest the existence of nonlinear sea surface temperature (SST)-rainfall relationships in the tropics and a strong influence of SST forcing on equatorial rainfall in the geographic vicinity of that forcing [*Hoerling et al.*, 1997]. To properly explain and ultimately predict this variability, it is important to disentangle, as far as possible, long range climatic phenomena from recent effects such as those possibly produced by deforestation and global warming.

[3] While the relationships among many climate and hydrological variables are decidedly nonlinear [*Jin et al.*, 2005], *linear* dependence measures are still being used as a matter of course to relate ENSO and inter-annual variability in river flows. These measures have ranged from linear correlation coefficients (CC) in the time domain [*Eltahir*, 1996; *Amarasekera et al.*, 1997; *Whitaker et al.*, 2001; *Anctil and Coulibaly*, 2004] to the cross-spectrum analysis [*Richey et al.*, 1989; *Wang and Eltahir*, 1999]. One of the reasons for using linear measures is that the inherent noise and periodicity in the observations together with short length of the available sample sizes make it difficult to use nonlinear approaches in climate and hydrology [*Tziperman et al.*, 1994; *An and Jin*, 2004; *Khan et al.*, 2005].

[4] The goal of this study is to investigate the nonlinear dependence between ENSO and the annual flow of some of the largest tropical and subtropical rivers, specifically the Nile, Amazon, Congo, Paraná and Ganges, through a measure based on the mutual information (MI). The results reveal a stronger extrabasinal connection between ENSO and river flows than the one suggested by linear analysis using linear regression (LR). This has significant impacts on scientific understanding and predictability as well as management of water and agricultural resources in vast, densely populated regions of the globe.

2. Data and Methodology

2.1. ENSO and River Flow Data

[5] ENSO events are associated with SST anomalies over the eastern and central equitorial Pacific Ocean. In this study, the ENSO index is defined in terms of the monthly SST variations from the long-term mean, averaged over the regions 2°–6°N, 90°–170°W; 2°N–6°S, 90°–180°W; and 6°–10°S, 110°–150°W of the Pacific Ocean. This dataset was published as a homogenized monthly series of the mean SST anomaly for the period 1872–1989 [*Wright*, 1989]. After 1989, the NINO 3.4 is used as the ENSO index because its geographical regions (5°N–5°S; 120°–170°W) are close to regions corresponding to the Wright SST.

[6] The monthly discharge data of the Nile River was measured at Aswan (lat. 24.1°N, long. 33°E) from 1873 to 1989. This integrated runoff comprises contributions from three major tributaries, specifically the White Nile, the Blue Nile, and the Atbara, and represents the majority of the Nile basin. The seasonal streamflow cycle of the Nile indicates that the minimum and maximum discharges are observed in April and September, respectively (Figure S1b in the auxiliary material).[1]

[7] The discharge data of the Amazon River was collected monthly from the Rio Negro stage at Manaus (lat. 3°S,

[1]Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.
[2]Department of Civil and Environmental Engineering, University of South Florida, Tampa, Florida, USA.
[3]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA.
[4]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

0094-8276/06/2006GL027941$05.00

[1]Auxiliary material data sets are available at ftp://ftp.agu.org/apend/gl/2006gl027941. Other auxiliary material files are in the HTML.

long. 60°W) over the period from 1903 to 1985. The integrated runoff at the Manaus gauge covers more than 3M km^2 of the Andean and western Amazon watershed [*Richey et al.*, 1989]. The seasonal streamflow cycle of the Amazon indicates that the minimum and maximum discharges are observed in November and June, respectively (Figure S1b).

[8] The Congo River discharge data was collected monthly from the river stage at Kinshasa, Zaire (lat. 4.3°S, long. 15.3°E) from 1905 to 1985. As the Congo River basin, covering approximately 3.8M km^2, is located around the equator, it experiences a marked semi-annual rainfall cycle which is associated with the north/south movement of the inter tropical convergence zone (ITCZ) across tropical Africa [*Todd and Washington*, 2004]. This is evident from the seasonal cycle of the Congo river indicating two peaks in May and December (maximum) and the lowest flow in August (Figure S1b).

[9] The Paraná River discharge data for the period 1904–1997 was collected monthly at Corrientes (lat. 27°S, long. 59°W) located downstream of the confluence of the Paraguay and the Paraná rivers. The seasonal cycle of the Paraná exhibits a single peak in February with a long recession and low discharge in September (Figure S1b).

[10] The monthly Ganges River discharge data was recorded over the period from 1934 to 1993 at the Hardinge Bridge in Bangladesh by the Bangladesh Water Development Board. It experiences the flood season from July to October, during which the average annual flow is 82% [*Whitaker et al.*, 2001]. The peak flow and low discharge of the Ganges are observed in August and April, respectively, as exhibited by the seasonal cycle (Figure S1b).

[11] During the year following the warm episodes of ENSO, the annual discharges of the Nile, Amazon, Congo and Ganges Rivers fall below their average annual discharge whereas the annual Paraná discharge is higher than the average annual discharge (Figure S1a). The runoff statistics give an idea about the discharge characteristics of the rivers (Table S1).

2.2. Mutual Information (MI)

[12] MI is a measure of statistical dependence among random variables which captures the full dependence structure, both linear and nonlinear. The concept of MI was originally developed in communication theory and has been applied to multiple domains over the last few decades [*Fraser and Swinney*, 1986; *Kraskov et al.*, 2004]. Considering two random variables X and Y, the MI, denoted by $I(X; Y)$, is defined as

$$I(X;Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y), \quad (1)$$

where $H(X)$ or $H(Y)$ is the marginal information entropy which measures the information content in a signal and $H(X, Y)$ is the joint information entropy which measures the information content in a joint system of X and Y. The MI between two random variables X and Y can also be defined as

$$I(X;Y) = \int_Y \int_X p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)} \, dx \, dy, \quad (2)$$

where $p_{XY}(x, y)$ is the joint probability density function (*pdf*) between X and Y, and $p_X(x)$ and $p_Y(y)$ are the marginal *pdfs*. The MI values range from 0 (independent) to ∞ (completely dependent). For a bivariate normal set (X, Y), the MI and the linear CC, denoted by ρ, are related as $I(X; Y) = -0.5 \log[1 - \rho(X, Y)^2]$ [*Joe*, 1989]. For comparing linear and nonlinear dependence measures, the MI-based nonlinear CC, i.e., λ, ranging from 0 to 1 is defined from the above relationship as

$$\hat{\lambda}(X,Y) = \sqrt{1 - \exp\left[-2\hat{I}(X;Y)\right]}, \quad (3)$$

where $\hat{\lambda}(X, Y)$ and $\hat{I}(X; Y)$ are the estimated nonlinear CC and MI between X and Y, respectively [*Joe*, 1989; *Granger and Lin*, 1994]. In addition, just as the mean squared errors (MSE) can be derived from LR, a lower bound of MI-based MSE, which is a measure of the predictability of Y based on the information content in X, can be estimated as

$$\widehat{MSE}(Y) \geq \frac{1}{2\pi e} \exp\left[2\left(\hat{H}(Y) - \hat{I}(X;Y)\right)\right], \quad (4)$$

where $\hat{H}(Y)$ is the estimated entropy of Y and $\hat{I}(X; Y)$ is the estimated MI between X and Y [*Brillinger*, 2004]. ANOVA-like interpretations have also been suggested for MI-based dependence [*Brillinger*, 2004]. *Cellucci et al.* [2005] compared MI-based dependence with traditional measures of dependence, such as Pearson linear correlation coefficient, Spearman rank order correlation, and Kendall's tau.

[13] The measures for linear (ρ) and nonlinear (λ) correlation quantify the strength of dependence among multiple variables (viz., ENSO and streamflow in this study). While the former quantifies the dependence purely in terms of the linear information content and the latter quantifies the complete (linear and nonlinear) information content, the two measures can be related in principle since they both capture the information contained in one variable about the other. The relationship between the two measures (ρ and λ) has been explained in detail by *Brillinger* [2004]. In more rigorous terms, the two measures can be compared quantitatively since they directly relate to the expected MSEs from predictions (see equation (4) for MSE from the MI-based dependence). The confidence bounds reflect the degree of belief in the two measures and hence can be compared as well. The definition of the nonlinear measure (λ) used here has been utilized by previous researchers [*Joe*, 1989; *Granger and Lin*, 1994; *Brillinger*, 2004] precisely because λ collapses to the linear measure (ρ) for the bivariate normal distribution (see equation (3)). We compare ρ from LR and λ obtained from first estimating the MI after fitting bivariate normal distribution to the data and then using equation (3). We test one simulation, i.e., chaotic (described in section 4.1), and one real data, i.e., dependence between ENSO and Nile River flow, and observe that λ obtained after fitting bivariate normal distribution is exactly similar to ρ for both cases (Figure S2). Finally, we would like to emphasize that the statements that compare linear and nonlinear correlation measures, while statistically valid,

need to be evaluated with care owing to issues pertaining to statistical estimation like bias-variance tradeoffs.

3. MI Estimation Methods

[14] The estimation of the MI requires the estimation of the joint and marginal *pdfs*, which, in turn, are frequently obtained from histogram and kernel density based estimators. Estimates of MI are consistent and asymptotically converge to the *true* or theoretical value when the data sets are relatively large and error-free. Since observations of river flows and the ENSO index are short and usually affected by various errors, it is important to assess various MI estimation methods for short and noisy data. Recently developed methodologies have been explored for estimating the MI, such as kernel density estimators (KDE) [*Moon et al.*, 1995], k-nearest neighbors (KNN) [*Kraskov et al.*, 2004], and Edgeworth approximation of differential entropy (Edgeworth) [*Hulle*, 2005].

3.1. Kernel Density Estimator (KDE)

[15] For any bivariate data set (X, Y) of size N, $\hat{I}(X; Y)$ is estimated as

$$\hat{I}(X;Y) = \frac{1}{N}\sum_{i=1}^{N}\log\frac{\hat{p}_{XY}(x_i,y_i)}{\hat{p}_X(x_i)\hat{p}_Y(y_i)}, \tag{5}$$

where $\hat{p}_{XY}(x_i, y_i)$ is the estimated joint *pdf*, and $\hat{p}_X(x_i)$ and $\hat{p}_Y(y_i)$ are the estimated marginal *pdfs* at (x_i, y_i).

[16] The multivariate kernel density estimator using a normal kernel is defined as

$$\hat{p}_X(\boldsymbol{x}) = \frac{1}{Nh^d}\sum_{i=1}^{N}\frac{1}{\sqrt{(2\pi)^d|\mathbf{S}|}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{x}_i)^T\mathbf{S}^{-1}(\boldsymbol{x}-\boldsymbol{x}_i)}{2h^2}\right), \tag{6}$$

where N is the number of data points; $\boldsymbol{x}$ and $\boldsymbol{x}_i$ are the d-dimensional vectors; $\mathbf{S}$ is the covariance matrix on the $\boldsymbol{x}_i$; $|\mathbf{S}|$ is the determinant of $\mathbf{S}$; and h is the kernel bandwidth also called the smoothing parameter [*Moon et al.*, 1995]. In this study, the smoothing parameter is chosen as the optimal Gaussian bandwidth for a normal kernel given as $h = [4/(d+2)]^{1/(d+4)} N^{-1/(d+4)}$. The MI estimates are obtained by first estimating $\hat{p}_X$, $\hat{p}_Y$, and $\hat{p}_{XY}$ from equation (6) and then plugging them in equation (5).

3.2. The k Nearest Neighbors (KNN)

[17] The MI between X and Y is estimated as

$$\hat{I}(X;Y) = \psi(k) - \frac{1}{k} - \frac{1}{N}\sum_{i=1}^{N}\left[\psi(n_x(i)) + \psi\big(n_y(i)\big)\right] + \psi(N), \tag{7}$$

where N and k are the number of data points and nearest neighbors, respectivley; if $\epsilon(i)/2$ is the distance between (x_i, y_i) and its kth neighbor, denoted by (kx_i, ky_i), and if $\epsilon_x(i)/2$ and $\epsilon_y(i)/2$ are given as $\|x_i - kx_i\|$ and $\|y_i - ky_i\|$, respectively, then $n_x(i)$ is the number of points x_j such that $\|x_i - x_j\| \le \epsilon_x(i)/2$; $n_y(i)$ can be calculated similarly; $\psi(x)$ is the digamma function, $\psi(x) = \Gamma(x)^{-1}d\Gamma(x)/dx$, which satisfies the relation $\psi(x+1) = \psi(x) + 1/x$, with $\psi(1) = -C$, where $C = 0.5772156649$ is the Euler-Mascheroni constant [*Kraskov et al.*, 2004]. This study chooses k as 3 since *Kraskov et al.* [2004] suggested $k > 1$ in order to reduce statistical errors and also indicated to avoid large values of k which lead to the increase of systematic errors.

3.3. Edgeworth Approximation of Differential Entropy (Edgeworth)

[18] Using Edgeworth expansion of the density $p(\boldsymbol{x})$, $\boldsymbol{x} = [x_1, .., x_d]$, the differential entropy is defined as

$$H(p) = H(\phi_p) - J(p) = H(\phi_p) - \frac{1}{12}\sum_{i=1}^{d}\left(\kappa^{i,i,i}\right)^2 - \frac{1}{4}\sum_{i,j=1,i\neq j}^{d}\left(\kappa^{i,i,j}\right)^2 - \frac{1}{72}\sum_{i,j,k=1,i<j<k}^{d}\left(\kappa^{i,j,k}\right)^2, \tag{8}$$

where d is the dimension of $\boldsymbol{x}$; $H(\phi_p) = 0.5 \log |\mathbf{S}| + \frac{d}{2}\log 2\pi + \frac{d}{2}$, where $\mathbf{S}$ is the covariance matrix, is the d-dimensional entropy of the best normal estimate, i.e., ϕ_p, with the same mean and covariance matrix as p; and κ is a standardized cumulant [*Hulle*, 2005]. In equation (8), $J(p)$ is called negentropy, which measures the distance to normal distribution. The MI is estimated by first estimating $\hat{H}(X)$, $\hat{H}(Y)$, and $\hat{H}(XY)$ from equation (8) and then plugging them in equation (1).

4. Analysis of Simulations

[19] We evaluate and compare MI estimation methods, i.e., KDE, KNN, and Edgeworth, using some simulations to find the best method for the real data analysis. Nonlinear CCs obtained from these methods are compared with linear and theoretical CCs using linear, nonlinear, and periodic functions, as well as the nonlinear Henon map, contaminated with different levels of artificial noise for small and large datasets. In this study, 50 and 100 points (comparable to the sizes of the geophysical data sets used in the study) and 1000 points are considered as short and long time series, respectively.

4.1. Details of the Simulated Data

[20] Case 1 (Linear): simple linear functions with Gaussian noise (ε) are used, such as $X \sim N(0, 1)$, Y: $y_i = x_i + \varepsilon_i$ for $i = 1, .., N$, where X is independent and identically distributed (iid) and $\varepsilon \sim N(0, \sigma_n)$ is *iid* and independent of X. Case 2 (Quadratic): simple quadratic functions with Gaussian noise are used, such as $X \sim N(0, 1)$, Y: $y_i = x_i^2 + \varepsilon_i$ for $i = 1, .., N$, where X and ε have the same meaning described above. Case 3 (Periodic): the periodical system with Gaussian noise is also analyzed, such as X, Y: $y_i = \sin(x_i) + \varepsilon_i$ for $i = 1, .., N$, where X is uniformly distributed between $-\pi$ to π and $\varepsilon \sim N(0, \sigma_n)$ is *iid* and independent of X. Case 4 (Chaotic): the Henon map, which exhibits chaotic behavior, is H_X: $H_{x_{i+1}} = 1 - \alpha H_{xi}^2 + Hy_i$, H_Y: $H_{y_{i+1}} = \beta H_{x_i}$ for $i = 1, .., N$, where $\alpha = 1.4$; $\beta = 0.3$; and $(H_{x_0}, H_{y_0}) = (0.0, 0.0)$. The Henon map with Gaussian noise is also analyzed, such as X: $x_i = H_{xi} + \varepsilon x_i$, Y: $y_i = H_{y_i} + \varepsilon y_i$ for $i = 1, .., N$, where $\varepsilon x \sim N(0, \sigma_{H_X})$ and $\varepsilon y \sim N(0, \sigma_{H_Y})$ are *iid* and independent of H_X and H_Y, respectively, and σ_{H_X} and σ_{H_Y} are

standard deviations of H_X and H_Y, respectively. The formulations for computing theoretical values of MI for cases 1, 2, and 3 are described in section 3.1 of Text S1 of the auxiliary material.

4.2. Conclusion From Simulations

[21] The simulations indicate that the presence of noise typically leads to an under-estimation of the *true* MI between the underlying nonlinear signals (see Text S1, section 3). As compared to KNN and Edgeworth, KDE is found to capture the underlying nonlinear dependence more consistently between two time series when they are short and noisy assuming such dependence exists (see Text S1, section 3). We also compare nonlinear dependence measures, such as KDE, KNN, and Edgeworth, with a rank-based dependence measure, i.e., Kendall's tau. From Kendall's tau, we observe a large negative bias in nonlinear dependence in the simulated data contaminated with noise (see Text S1, section 4). Thus in this study LR and KDE approaches have been consistently used to estimate and compare linear and nonlinear CCs respectively.

5. Real Data Analysis

[22] This study assumes that the seasonal cycle for a particular year consists of 12 months starting with the month having the lowest average discharge. It also assumes that long-term flow variability due to ENSO can be captured in the annual flow, which, in turn, is defined as the integrated streamflow of the seasonal cycle. Here eight quarterly ENSO indices, specifically three quarters just before the seasonal cycle, four quarters corresponding to the seasonal cycle, and one quarter just after the seasonal cycle, are derived from quarterly averages of mean monthly SST anomalies. The bivariate normal and kernel density between the quarterly ENSO indices and the annual flow of the Nile, Amazon, Congo, Paraná, and Ganges rivers are estimated and plotted (Figures S3–S7). Linear and nonlinear CCs between the ENSO index and the annual flow of the Nile, Amazon, Congo, Paraná, and Ganges Rivers are obtained using LR and KDE, respectively (Figure 1). The bias-corrected CCs and their 90% confidence bounds are estimated using jackknifing.

[23] The jackknife is used to estimate the bias-corrected λ and ρ and their standard errors using KDE and LR, respectively. The technique is described below for λ and is the same for ρ. In the case of real data analysis, the total number of observations (N) varies from 60 to 117. If d observations for jackknifing are left out and $\sqrt{N} < d < N$, the number of jackknife samples, given by $\binom{N}{d}$ [*Efron and Tibshirani*, 1993], is large. So 100 samples of size $0.8N$ are used for the analysis. $\hat{\lambda}^*(.)$ is the mean of jackknife replications leaving out $d = 0.2N$ observations. The bias is given as $\widehat{bias} = \hat{\lambda}^*(.) - \hat{\lambda}$, where $\hat{\lambda}$ is the original nonlinear CC between the annual flow and ENSO considering all N observations. The bias-corrected estimator, $\overline{\lambda}$, is given as $\overline{\lambda} = \hat{\lambda} - \widehat{bias}$. The lower and upper bounds of 90% confidence bounds are given by 5% and 95% quantiles of 100 jackknife samples of size $0.8N$, respectively.

[24] The prediction accuracies, in terms of MSEs, of annual river flows based on ENSO are also estimated and compared using LR and KDE approaches (Tables S2–S6).

5.1. Description of Results

[25] Linear CCs between river flows and some quarters of the ENSO index, such as, all quarters of the ENSO index and the Nile flow, quarter 2 to quarter 6 of the ENSO index and the Amazon flow, quarter 1 to quarter 7 of the ENSO index and the Congo flow, and all quarters of the ENSO index and the Ganges flow, are negative. Since nonlinear CCs obtained from KDE, KNN, and Edgeworth do not have directionality, the absolute values of linear CCs are considered and plotted. The MI-based nonlinear dependence measure, i.e. KDE, generate higher CCs and lower MSEs as compared to linear dependence measure, i.e. LR, which shows that KDE captures more extrabasinal connection between ENSO and river flows in the tropical and subtropical regions of the world as compared to LR (Tables 1 and S2–S6). The percentage variation in the annual flow of rivers associated with ENSO are calculated as the square of CCs. KDE suggests an increase of around 20–70% in the extrabasinal connection between ENSO and river flows over those suggested by LR (Figure 1 and Table 1). In the case of Nile, 90% confidence bounds of linear and nonlinear CCs are well separated for 5 quarters including quarter with the highest nonlinear CC indicating that KDE captures greater dependence between ENSO and the annual flow compared to LR (Figure 1a). KDE suggests greater dependence between the Congo flow and ENSO since 90% confidence bounds of linear and nonlinear CCs are well separated for all quarters except the first quarter (Figure 1c). In the case of Amazon, Paraná, and Ganges, 90% confidence bounds of linear and nonlinear CCs overlap for all those quarters which have higher linear CCs but for other quarters the bounds are well separated (Figures 1b, 1d, and 1e). This indicates that both KDE and LR capture nothing more than the linear dependence for some quarters based on 90% confidence bounds. However, there is an increase in the bias-corrected CCs from KDE as compared to LR for the Amazon, Paraná, and Ganges Rivers which suggests a stronger extrabasinal connection between ENSO and the annual flow of these rivers, however with less than 90% confidence (Figures 1b, 1d, and 1e). When linear CCs are close to zero, the large difference between linear and nonlinear CCs should be interpreted with caution because of an artifact of equation (3) which scales nonlinear CCs exponentially with MI (Equation (3) and Figures 1c and 1e).

5.2. Conclusion From the Analysis

[26] The results with the real data reported here suggest that there exists a nonlinear extrabasinal connection between ENSO and river flows in the topics and subtropics. This study also shows an appreciable increase in the variation of annual river flows linked to ENSO using nonlinear relationship measure as compared to linear measures. Hence, these results indicate additional predictability in the ENSO-streamflow extrabasinal connection when MI-based approaches are used, as compared to linear approaches used by researchers till date. The additional dependence captured by the MI-based nonlinear CCs may be useful for developing more accurate and longer streamflow models. This can, in turn, help in water resources management (e.g., reservoirs

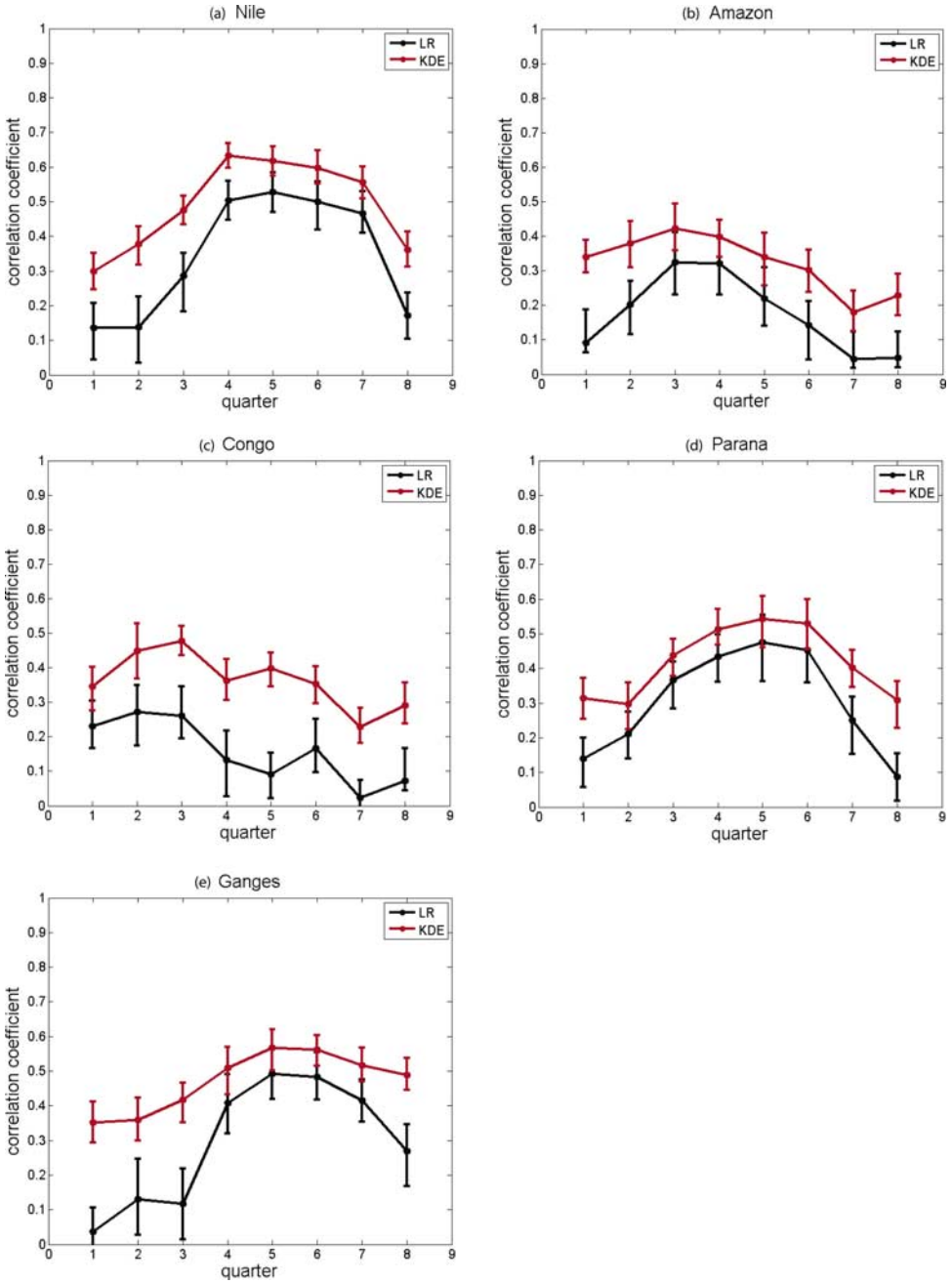

Figure 1. Nonlinear and linear CCs with their 90% confidence bounds between ENSO and annual river flows of (a) Nile, (b) Amazon, (c) Congo, (d) Paraná, and (e) Ganges using KDE and LR approaches, respectively. The bias-corrected estimates, $(\overline{\lambda}, \overline{\rho})$, plotted as solid dots are estimated as $2(\hat{\lambda}, \hat{\rho}) - (\hat{\lambda}^*(.), \hat{\rho}^*(.))$, where $(\hat{\lambda}, \hat{\rho})$ are the original nonlinear and linear CCs between the annual flow and ENSO, respectively, considering all N observations. $(\hat{\lambda}^*(.), \hat{\rho}^*(.))$ is the mean of 100 jackknife replications of size $0.8N$. The 90% confidence bounds are given by 5% and 95% quantiles of 100 jackknife replications of size $0.8N$.

Table 1. Variation in the Annual Flow of Rivers Associated With ENSO[a]

River	Previous Studies	Linear CC	Nonlinear CC
Nile	25% [SON] [*Eltahir*, 1996]	28% [ASO]	40% [MJJ]
Amazon	10% [D$^-$JF] [*Amarasekera et al.*, 1997]	11% [A$^-$S$^-$O$^-$]	18% [A$^-$S$^-$O$^-$]
Congo	10% [MAM] [*Amarasekera et al.*, 1997]	7% [F$^-$M$^-$A$^-$]	23% [M$^-$J$^-$J$^-$]
Paraná	19% [D$^-$JF] [*Amarasekera et al.*, 1997]	23% [DJF]	29% [DJF]
Ganges	29% [JJA] [*Whitaker et al.*, 2001]	24% [JAS]	32% [JAS]

[a]Linear and nonlinear CCs are estimated using LR and KDE, respectively. Months in a quarter are given in brackets. The month preceding the seasonal cycle is indicated by a negative sign following a month.

and dams for flood control, power generation, drought mitigation and preparedness for water supply).

6. Discussion

[27] Streamflow series may reflect monotonic trends related to anthropogenic factors, which may include diversions, consumptions and flow regulations within the basin, in addition to possible impacts of climate change. An estimation of the likely magnitudes, as well as qualitative assessment of the evidence, of such changes may need to be performed on a case by case basis for each basin. Just as an example, three of the co-authors of this paper performed qualitative investigations for streamflows of two rivers within the United States [*Khan et al.*, 2005]. Our investigations demonstrated that thorough studies can be time-consuming and intensive, hence such efforts are left as areas of future research. Discussions regarding the specific datasets utilized in this paper can be found from the data sources as well as within the previous studies that have utilized these datasets [*Richey et al.*, 1989; *Eltahir*, 1996; *Amarasekera et al.*, 1997; *Whitaker et al.*, 2001]. We would like to note that accounting for all the known trends, if possible, may have further impact on the ENSO to streamflow connection. Thus, it is likely that the ENSO-streamflow extrabasinal connection is actually even higher than estimated if such trends were to be accounted for. Conversely, it is possible that some extremes are highlighted in the anthropogenic basin flow trends which tend to overemphasize the ENSO connection. On the other hand, an argument can perhaps be made that such situations are not relevant to the point of this paper since the influence of anthropogenic or other trends will be reflected in both the linear and nonlinear measures of dependence. However, while making a priori statements may not be justified, it is likely that some of the trends will be nonlinear and that the nonlinear measures may be potentially more susceptible to the presence of outliers.

[28] Although ENSO has a direct influence on rainfall anomalies over the tropical and subtropical regions, only a portion of the variation in the annual flow of rivers located in these regions is associated with ENSO events. This is most likely caused by the complex ocean-land-atmosphere interactions, rainfall-runoff relationships, and anthropogenic influences, compounded by the fact that the available data may be noisy, incomplete or corrupted.

[29] In recent decades, economic, population and geo-political pressures have resulted in significant changes in land-use patterns that may alter the land-atmosphere-water cycle in the tropics and subtropics. These changes in the water cycle can, in turn, impact regional precipitation, water vapor flux, and surface water flows, causing regional as well as global shifts in seasonal-to-interannual atmospheric phenomena. A better understanding and quantification of the relationship between ENSO and river discharges can help scientists and policy makers understand and get prepared for the changes in river discharge patterns, in addition to attributing such changes to natural or anthropogenic drivers.

[30] **Acknowledgments.** This research was partially funded by the SEED money funds of the Laboratory Directed Research and Development Program of the Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U.S. DOE under contract DE-AC05-00OR22725 (Project title: "Multivariate dependence in climate extremes"; PI: Auroop R. Ganguly). The authors thank the following scientists who provided or pointed us to the data used in this research: Guiling Wang of the University of Connecticut, Earle Williams of the Massachusetts Institute of Technology (MIT), Shafiqul Islam of Tufts University, Norberto O. Garcia of the Universidad Nacional del Litoral, Argentina, and Carlos Nobre of the Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, Brazil. We are thankful to Alexander Kraskov for providing us with KNN-based MI code. Helpful comments from Rafael Bras of MIT, Thomas Wilbanks, Gabriel Kuhn, Jim Nutaro, Alexander Sorokine of ORNL, and Bellie Sivakumar of the University of California, Davis, are gratefully acknowledged. The authors are thankful to an anonymous reviewer for helpful suggestions which significantly improved the quality of the paper. In addition, the second author acknowledges the support provided by the *SensorNet*® research program at ORNL and his courtesy faculty affiliation at the University of South Florida.

References

Amarasekera, K. N., R. F. Lee, E. R. Williams, and E. A. B. Eltahir (1997), ENSO and the natural variability in the flow of tropical rivers, *J. Hydrol.*, *200*, 24–39.

An, S.-I., and F.-F. Jin (2004), Nonlinearity and symmetry of ENSO, *J. Clim.*, *17*, 2399–2412.

Anctil, F., and P. Coulibaly (2004), Wavelet analysis of the interannual variability in southern Québec streamflow, *J. Clim.*, *17*, 163–173.

Brillinger, D. R. (2004), Some data analyses using mutual information, *Braz. J. Probab. Stat.*, *18*, 163–183.

Cellucci, C. J., A. M. Albano, and P. E. Rapp (2005), Statistical validation of mutual information calculations: Comparisons of alternative numerical algorithms, *Phys. Rev. E*, *71*, 066208.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, CRC, Boca Raton, Fla.

Eltahir, E. A. B. (1996), El Niño and the natural variability in the flow of the Nile River, *Water Resour. Res.*, *32*(1), 131–137.

Fraser, A. M., and H. L. Swinney (1986), Independent coordinates for strange attractors from mutual information, *Phys. Rev. A*, *33*(2), 1134–1140.

Granger, C., and J. Lin (1994), Using the mutual information coefficients to identify lags in nonlinear models, *J. Time Series Anal.*, *15*(4), 371–384.

Hoerling, M. P., A. Kumar, and M. Zhong (1997), El Niño, La Niña, and the nonlinearity of their teleconnections, *J. Clim.*, *10*, 1769–1786.

Hulle, M. M. V. (2005), Edgeworth approximation of multivariate differential entropy, *Neural Comput.*, *17*, 1903–1910.

Jin, Y. H., A. Kawamura, K. Jinno, and R. Berndtsson (2005), Nonlinear multivariable analysis of SOI and local precipitation and temperature, *Nonlinear Processes Geophys.*, *12*, 67–74.

Joe, H. (1989), Relative entropy measures of multivariate dependence, *J. Am. Stat. Assoc.*, *84*(405), 157–164.

Khan, S., A. R. Ganguly, and S. Saigal (2005), Detection and predictive modeling of chaos in finite hydrological time series, *Nonlinear Processes Geophys.*, *12*, 41–53.

Kraskov, A., H. Stögbauer, and P. Grassberger (2004), Estimating mutual information, *Phys. Rev. E*, *69*, 066138.

Moon, Y., B. Rajagopalan, and U. Lall (1995), Estimation of mutual information using kernel density estimators, *Phys. Rev. E*, *52*(3), 2318–2321.

Richey, J. E., C. Nobre, and C. Deser (1989), Amazon River discharge and climate variability: 1903–1985, *Science*, *246*, 101–103.

Todd, M. C., and R. Washington (2004), Climate variability in central equatorial Africa: Influence from the Atlantic sector, *Geophys. Res. Lett.*, *31*, L23202, doi:10.1029/2004GL020975.

Tziperman, E., L. Stone, M. A. Cane, and S. Zebiak (1994), El Niño chaos: Overlapping of resonances between the seasonal cycle and the Pacific ocean-atmosphere oscillator, *Science*, *264*, 72–74.

Wang, G., and E. Eltahir (1999), Use of ENSO information in medium- and long-range forecasting of the Nile floods, *J. Clim.*, *12*, 1726–1737.

Whitaker, D. W., S. A. Wasimi, and S. Islam (2001), The El Niño–Southern Oscillation and long-range forecasting of flows in the Ganges, *Int. J. Climatol.*, *21*, 77–87.

Wright, P. B. (1989), Homogenized long-period Southern Oscillation indices, *Int. J. Climatol.*, *9*, 33–54.

S. Bandyopadhyay, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

D. J. Erickson III and G. Ostrouchov, Computer Science and Mathematics Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA.

A. R. Ganguly, S. Khan, and V. Protopopescu, Computational Sciences and Engineering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA. (gangulyar@ornl.gov)

S. Saigal, Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL 33620, USA.